Perception Analysis for University of San Carlos (USC) as an Educational Institution using Web Mining and Multinomial Naïve Bayes Algorithm

Angie M. Ceniza
University of San Carlos
School of Arts and Sciences
Department of Computer and Information Sciences

Christian V. Maderazo
University of San Carlos
School of Arts and Sciences
Department of Computer and Information Sciences

Mary Jane G. Sabellano
University of San Carlos
School of Arts and Sciences
Department of Computer and Information Sciences

Abstract:- **The use of blogs, forums, and other forms of media over the web in order to express one's perception is tremendously increasing that either endorses or criticizes the performance of an entity or organization. The aim of this research is to determine the public's perception of University of San Carlos (USC) as an academic institution specifically in the classification of positive, negative or neutral polarity through Sentiment Analysis. In this paper we explore the use of Web Mining as the process of harvesting hyperlink structure and Multinomial Naïve Bayes Algorithm to evaluate web content as positive, negative or neutral perceptions. The experiment results show that the approach used in the research achieves 52.92% precision and 92.88% recall.**

Keywords: Perceptions, Sentiment Analysis, Web Mining, Multinomial Naïve Bayes Algorithm

1. INTRODUCTION

Perception analysis is a key in knowing one's reputation. In an educational institution, it is an advantage to know their reputation through public sentiments. According to Pang and Lee [31] "What other people think" is an important piece of information during the decision making process. Administrators will be able to evaluate the performance of the university as provider of quality education. Sentiment analysis is a computational treatment of people's opinions, attitudes and emotions towards an entity [25]. It helps individuals and organization in making decisions [31]. In gathering sentiments, web mining is one of the most conducive techniques since most of the people nowadays use different forms of media over the web as their means of expressing ideas. Researchers make use of this technique in gathering the information needed. Web data mining technique [23] aims to discover useful information or knowledge from the Web hyperlink structure, page content, and usage data. The Web mining process is similar to the data mining process. The difference is usually in the data collection. In traditional data mining, the data is often already collected and stored in a data warehouse. For Web mining, data collection can be a substantial task, especially for Web structure and content mining, which involves crawling a large number of target Webpages. The researchers gather all related information to University of San Carlos – Cebu into the web. Those "bag-of-words" that will be collected will be analyzed and evaluated by Multinomial Naïve Bayes classifier [35]. Medhat, Hassan and Korashy [25] identified the algorithm as one of the most frequently used for solving Sentiment classification problem. Bhadane, Dalal and Doshi [1] considered it as commonly used algorithm. For this reason, researchers decided to apply this algorithm in identifying the public's perceptions of University of San Carlos as an academic institution.

2. REVIEW OF RELATED LITERATURE

These are several lines of related work which are reviewed in this section.

2.1 Perception Analysis

Most users publish personal messages about people, product, events and interest. These opinionated messages are available for study. The classification of user's sentiments has been recognized as a significant research area. Sentiment Analysis has emerged as a rapidly expanding field of application and research in the area of information retrieval [6]. In the published work of Pang and Lee [29] they conducted a study of sentiment analysis that seeks to identify the viewpoint(s) underlying a text span. In the research of Pathak, Mane, Srivastava and Contractor [31], they study about perception of knowledge in an organization through an email network. Iskender and Bati [13] conducted Sentiment Analysis in the context of

Entrepreneurship and Innovativeness to obtain two university rankings. Kim and Hovy [18] identified sentiments as a challenging problem. Kandiko and Mawer [15] integrated the expectations and perceptions of the quality of the learning experience and the academic standards of their chosen programs of study. Ravindran and Kalpana [34] conducted a study that would assess management education students' expectation, perception, and satisfaction of services they experienced. In [22] the authors developed an adaptive sentiment analysis models for online reviews.

2.2 Web Mining

Web mining is the application of data mining and is used for the same task [39]. Kosala and Blockeel[19] published about the importance of web mining research in several research communities as well as sub-areas of machine learning and natural language processing. Wang [44] believed that web mining is consist of pre-processing, pattern discovery and pattern analysis. Kosala and Blockeel[19] conducted survey that point some confusion regarding the usage of the term Web Mining and suggest three Web Mining categories. Guidici and Catelo [11] considered a model-based approach to web mining. Srivastava, Cooleyz, Deshpande and Tan[41] identified the detailed taxonomy of Web Usage Mining. In the study of [27] the user preference is automatically learned from web usage data by using data mining techniques. In [4] the authors presented the details of preprocessing tasks that are necessary for peforming Web Usage Mining, the application of data mining and knowledge discovery tech niques to WWW server access log. In the study of [16] the labeling process could benefit from employment of web mining and information extraction techniques in combination with flexible methods of web-based information management. Furnkanz[10] provided a brief overview of web mining techniques and research areas such as hypertext classification, wrapper induction, recommender system and web usage mining.

2.3 Multinomial Naïve Bayes

Multinomial Naïve Bayes is a popular method for document classification due to its computational efficiency and relatively good predictive performance [9]. Dhande and Patnaik[7] stated that sentiment classification is an effective tool in classifying reviews of user using positive or negative from textual information alone. Shimodaria[38] identified that an effective way of classifying text by their contents is text classification especially in classifying email messages into spam or non-spam. Kibriya, Frankm, Pfahringer and Holmes [17] believed that the use of Multinomial Naïve Bayes is an effective tool in classifying text and by adding a locally weighted learning. The authors in [3] used two classical classification methods, Naïve Bayes and SVM, to test whether the selected sentiment are useful for different classification methods. In [14] the authors build multinomial event models for SAGE data classification based on Naïve Bayes classifier. From the analysis of the experimental human brain and breast SAGE data, the Multinomial Naïve Bayes model can achieve very high classification accuracy. In [45] authors proposed Naïve Bayes Tree hybrid algorithm, which deploys a Naïve Bayes classifier on each leaf node of the built decision tree and has demonstrated remarkable classification performance. Shan and Banerjee [37] proposed a family of mixed-membership Naïve Bayes models that extends the popular Naïve Bayes models to work with sparse observation by marginalizing over all missing features. Wong [46] introduced generalized Dirichlet priors to preserve the prediction accuracy and computational efficiency of Naïve Bayes classifier. The authors in [28] introduced the combination of Expectation-Maximization and Naïve Bayes classifier to reduce classification error. In [5] the authors proposed an extension of the Naïve Bayes classifier which accounts for biological heterogeneity in a probabilistic framework. Maruyama [24] used the classifier on yeast data sets with higher predictability performance. Frank and Bouckaert [9] applied Multinomial Naïve Bayes to unbalanced datasets.

2.4 Opinion Lexicon

Opinion words are used in many sentiment classification research. Positive opinion words are used to express some desired states, while negative opinion words are used to express some undesired states [25]. The most commonly used lexicon sources are WordNet and Opinion Lexicon which exists in languages other than English. Building resources, used in Sentiment Analysis tasks, is still needed for many natural language. Hu and Lui[12][21] maintains and freely distributes a sentiment lexicon consisting of list of strings. In the study of [43] two experiments used a forty-word lexicon. Flekova, Ruppert, Pietro[8] uses open lexicon in identifying frequent bigrams where a polar word switches polarity. Cruz, Troyano, Ortega and Enriquez [2] computed the semantic orientation (positive and negative evaluative implications) of certain opinion expressions using opinion lexicon. Souza, Vieira, Busetti, Chishman and Alves [40] concluded in their research that Opinion Lexicons are linguistic resources annotated with semantic orientation of terms (positive and negative) and are important for opinion mining tasks. In [26] WordNet is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. Pustejovsky [32] provided a method for the decomposition of lexical categories and outlined a theory of lexical semantics.

3. METHODOLOGY

In this section, we describe the methods used to determine public perceptions about University of San Carlos as an academic institution. Figure 1.0 shows the processing flow of the approach.

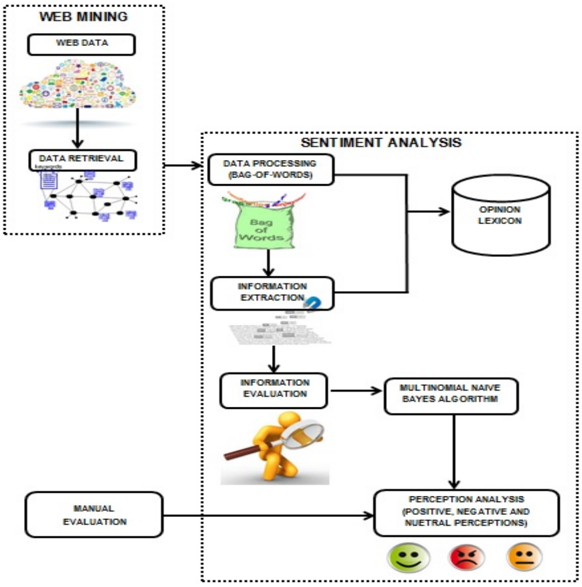

Figure 1.0 The processing flow of the approach

3.1 Web Mining

This research makes use of web mining in retrieving the information needed. Web mining [42] is the application of data mining techniques to extract knowledge from Web data, including Web documents, hyperlinks between documents, usage logs of web sites, etc. According to Liu [23], web mining aim to discover useful information or knowledge from the web hyperlink structure, page content, and usage data. The harvesting technique uses Link Klipper Chrome extension tool. The researchers make use of Google [42] as the search engine using the search keywords "University of San Carlos Cebu City". Google.com was able to harvest 654 links, ask.com is less than a 100 links, and bing.com is less that 250 links while yahoo.com was not able to harvest links. Web structured mining and web content mining is being applied in gathering the information needed in this research. Web Structure Mining is the process of discovering structure information from the web. The Web Content Mining is the process of extracting useful information from the contents of Web documents. Content data corresponds to the collection of facts a Web page was designed to convey to the users. Researchers extracted all the contents of 654 links into useful information that will be compared to an Opinion Lexicon.

3.2 Bag of Words Model

In this research, a document-level sentiment analysis classification was used. It considers the whole web content for each harvested link as a basic information unit. The document is represented as a bag of words model. The model knows which words are included in the document and its frequency count, and ignores the position of the word in the document. The bag of words model is used more often because of its simplicity for the classification process. In bag of words model, the occurrence of each word will be classified as positive and negative word through the use of Opinion Lexicon[12][21]. Single words from a document or n-gram (two words) from a document in sequential order are chosen as feature. In this research two words refers to words with negation word, such as "no", "not", "none", "nothing", "nowhere", "neither", "nobody", "hardly", "scarcely", "barely". This research makes use of Multinomial Naïve Bayes algorithm in sentiment classification.

3.3 Opinion Lexicon

Opinion Lexicon [17][18] is a list of identified positive and negative (a – z) words . There are 2,006 positive words and 4, 783 negative words. It includes mis-spelling words, morphological variants, slang and social-media mark-up. These words serve as trainings sets that enable us to identify positive and negative sentiments in a certain website.

Here are some of the positive and negative words found in the Opinion Lexicon.

Positive Words
{accessible, acclaim,
accolade, creative, dazzle,
decent, ease, educated,
elegant, fast-growing, fast-paced,
first-in-class, gentle, glow,
hearten, intelligent, joyful,
jubilate, kindness,
lucky, lucrative, luminous,
magical, marvel, non-violent, openly, pamper,
proven, proud,
reputation,
recommendation, satisfied,
secure, tender, thrive,
unrestricted, unreal,
valuable,
vibrant, well-educated,
well-managed, wholesome,
worth-while, zeal, zenith}

Negative Words
{ abruptly, abscond,
absence, absent-minded,
barren, baseless, bash,
bastard, cold, complain,
conceited, desperate,
destroy, devastated,
disadvantage,
entanglement,
fell, fibber, get-rich, grind,
halfheartedly, ill-advised,
ill-conceived,
ill-defined, jobless, joke,
lamentably, maliciously,
maltreatment, nightmare,
obstructed, outrageous,
paralize, parasite,
tattered, temper, terrorism,
unachievable, unaffordable,
unappealing,
venomous, vibrate,
watered-down, wayward,
yawn, zap}

3.4 Multinomial Naïve Bayes Algorithm

Multinomial Naïve Bayes (MNB) Algorithm [33] is used to characterize text documents in terms of frequency (tf,(t,d)). It is a popular method for document classification due to its computational efficiency and relatively good predictive performance [9]. A classification technique that performs well in many domains [36]. It is one of the top 10 algorithms because of its simplicity, efficiency and interpretability [45]. The term frequency is typically defined as the number of times a given term t (i.e., word or token) appears in a document d (this approach is sometimes also called raw frequency). In practice, the term frequency is often normalized by dividing the raw term frequency by the document length.

$$normalized\ term\ frequency = \frac{tf(t,d)}{n_d} \quad (1)$$

where:

- $tf(t,d)$: Raw term frequency (the count of term t in document d).
- n_d : The total number of terms in the document.

The term frequencies can then be used to compute the maximum-likelihood estimate based on the training data to estimate the class-conditional probabilities in the multinomial model:

$$\hat{P}(x_i|w_j) = \frac{\sum tf\,(x_{i,}\, d \in w_j) + \alpha}{\sum N_{d \in wj} + \propto .V} \quad (2)$$

where:

- x_i : A word from the feature vector x of a particular sample.
- $\sum tf(x_{i,}\, d\ \in\ w_j)$: The sum of raw term frequencies of word xi from all documents in the training sample that belong to class w_j.
- $\sum N_{d\ \in wj}$: The sum of all term frequencies in the training dataset for class w_j.
- α : An additive smoothing parameter (α=1 for Laplace smoothing).
- V : The size of the vocabulary (number of different words in the training set).

The class-conditional probability of encountering the text x can be calculated as the product from the likelihoods of the individual words (under the naive assumption of conditional independence).

$$P(x|w_j) = P(x_1|w_j) \cdot P(x_2|w_j) \cdot \ldots \cdot P(x_n|w_j) = \prod_{i=1}^{m} P\,(x_i|w_j) \quad (3)$$

In this research, the number of positive and negative words in every link (web page) being harvested will be counted using this MNB algorithm. Researchers decided to use this algorithm because according to Raschka[33] , MNB Algorithm is an effective algorithm in text categorization in counting "how often the words would occurs in the document" or "number of times outcome number x_i is observed over the end trial". Compared to Multi-Variate Bernoulli Naïve Bayes Algorithm that would classify text based on the occurrence (1 or 0's) - "words occurs in the document" or "words doesn't occur in the document". Table 1.0 shows some of the harvested links, the number of positive words, number of negative words and its sentiment classification.

Table 1.0 Harvested links with the number of positive and negative words

	Links	No. of Positive Words	No. of Negative Words	Result
1	https://www.linkedin.com/title/guidance-counselor-at-university-of-san-carlos	65	1	Positive
2	http://www.finduniversity.ph/education-schools/central-visayas/cebu-city/masters-programs/	31	0	Positive
3	http://www.livinginthephilippines.com/live-and-retire/schools/781-university-of-san-carlos	45	8	Positive
4	http://www.tripadvisor.com/HotelsNear-g294261-d7889978-University_of_San_Carlos-Cebu_Island_Visayas.html	81	8	Positive
5	http://www.tripadvisor.co.uk/HotelsNear-g294261-d7889978-University_of_San_Carlos-Cebu_Island_Visayas.html	77	9	Positive
6	http://www.university-directory.eu/Philippines/University-of-San-Carlos.html	4	2	Negative
7	http://www.filipinoscholar.com/200709-university-of-san-carlos-cebu-city-scholarships.html/	7	2	Neutral
8	http://www.tripadvisor.ie/HotelsNear-g294261-d7889978-University_of_San_Carlos-Cebu_Island_Visayas.html	72	8	Positive
9	http://uscedu.blogspot.com/	2	1	Negative
10	http://www.tripadvisor.co.uk/HotelsNear-g294261-d7889978-University_of_San_Carlos-Cebu_Island_Visayas.html	78	9	Positive

The graph in Figure 2.0 shows result of the perception analysis for University of San Carlos as an academic institution. There were 17% of the links were evaluated to have a negative perception, 76% were positive perception and 7% were neutral perceptions.

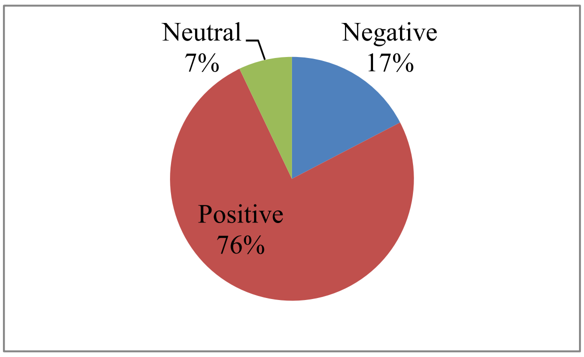

Figure 2.0 Public Perception for University of San Carlos as an academic institution

4. RESULTS

The measurement of performance evaluation is done by using precision and recall of the sentiment prediction.

$$Precision = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ documents\ retrieved} \quad (4)$$

$$Recall = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ relevant\ documents} \quad (5)$$

The researchers harvested 654 links related to University of San Carlos Cebu City. Each links were evaluated with the occurrence of positive and negative words. The researchers have found out that out of 654 harvested links there were 13 links considered as “forbidden links” and 290 links that produced “invalid perceptions” because they were no positive or negative perceptions being identified. In other words, there were only 351 links that enable the researchers to identify public perceptions to University of San Carlos as an academic institution. In the 351 links, 61 links have negative perceptions, 25 links with neutral perceptions and 265 links with positive perceptions. The experiment results show that the approach used in the research achieves 52.92% precision and 92.88% recall.

5. CONCLUSIONS

In this research, information gathering and analysis over the internet become so important in providing the researchers the public’s perception for University of San Carlos as an academic institution. We implemented a set of techniques using Web mining to harvest needed information and the use of Bag of Words model, Opinion Lexicon and Multinomial Naïve Bayes algorithm in the sentiment classification. Experiments results show that the system works at a good level, giving a relatively accurate representation of public’s perception to University of San Carlos. We aim to extend the system to harvest social networking post, consider using multiple words in the feature extraction and explore the use of hybrid sentiment classification algorithm.

6. ACKNOWLEDGEMENT

The researchers wish to extend our outmost gratitude to Department of Computer and Information Sciences for giving the opportunity to conduct this research. And, most especially to God Almighty, for giving us the wisdom in pursuing this research.

7. REFERENCES

[1] BHADANE, C., DALAL, H., AND DOSHI, H. Sentiment analysis: Measuring opinions. *Procedia Computer Science* 45, (2015), 808-814.

[2] CRUZ, F. L., TROYANO, J. A., ORTEGA, F. J., AND ENRÍQUEZ, F. Automatic expansion of feature-level opinion lexicons. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* (June 2011).

[3] CAO, D., JI, R., LIN, D., AND LI, S. Visual sentiment topic model based microblog image sentiment analysis. *Multimedia Tools and Applications*, (2014), 1-14.

[4] COOLEY, R., MOBASHER, B., & SRIVASTAVA, J. Data preparation for mining world wide web browsing patterns." *Knowledge and information systems* 1, 1 (1999), 5-32.

[5] DEMICHELIS, F., MAGNI,P., PIERGIORGI, P., RUBIN, M., AND BELLAZZI, R. A hierarchical Naïve Bayes Model for handling sample heterogeneity in classification problems: an application to tissue microarrays. BMC Bioinformatics 7, (December 2006), 514.

[6] DEVITT, A., AND AHMAD, K. Is there a language of sentiment? An analysis of lexical resources for sentiment analysis. *Lang Resources & Evaluation* 47, (2013), 475-511.

[7] DHANDE, L., AND PATNAIK, G. Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier. *International Journal of Emerging Trends & Technology in Computer Science* 3, 4 (July-August 2014), 313-320.

[8] FLEKOVA, L., RUPPERT, E., AND PREOTIU-PIETRO, D. Analysing domain suitability of a sentiment lexicon by identifying distributionally bipolar words. In *6th Workshop on Computational Approaches to Subjectivity, Sentiment And Social Media Analysis WASSA* (2015).

[9] FRANK, E., AND BOUCKAERT, R. Naive bayes for text classification with unbalanced classes. In *PKDD'06 Proceedings of the 10th European conference on Principle and Practice of Knowledge Discovery in Databases* (Berlin, Germany, 2006).

[10] Fürnkranz, Johannes. Web mining. Data mining and knowledge discovery handbook. Springer US, 2005.

[11] GIUDICI, PAOLO, AND ROBERT CASTELO. Association models for web mining. *Data mining and knowledge discovery* 5, 3 (2001), 183-196.

[12] HU, M., AND LIU, B. Mining and summarizing customer reviews. In *KDD ’04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, New York, NY, USA, August 2004).

[13] ISKENDER, E., AND BATI, G. B. Comparing Turkish Universities Entrepreneurship and Innovativeness Index’s Rankings with Sentiment Analysis Results on Social Media. *Procedia – Social and Behavioral Sciences* 195, (2015), 1543-1552.

[14] JIN, X., ZHOU, W., AND BIE, R. Multinomial event Bayesian modeling for SAGE data classification. *Computational Statistics* 22, 1 (April 2007), 133-143.

[15] KANDIKO, C. B., AND MAWER, M. Student expectations and perceptions of Higher Education. *London: King’s Learning Institute* , 2013.

[16] KARKALETSIS, V., STAMATAKIS, K., KARAMPIPERIS, P., LABSKÝ, M., RUZICKA, M., SVÁTEK, V., ... AND VILLAROEL GONZALES, D. Management of medical website quality labels via web mining. *Data Mining and Medical Knowledge Management: Cases and Applications. USA: IGI Global Inc,* (2009), 206-226.

[17] KIBRIYA, A. M., FRANK, E., PFARINGER, B., AND HOLMES, G. Multinomial Naive Bayes for Text Categorization Revisited. In *AI 2004: Advances in Artificial Intelligence (*Springer Berlin Heidelberg, 2005).

[18] KIM, S., AND HOVY, E. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics* (Association for Computational Linguistics, 2004).

[19] KOSALA, R., AND BLOCKEEL, H. Web Mining Research: A Survey. *ACM Sigkdd Explorations Letter* 2, 1 (200), 1-15.

[20] KUO, Y., FU, M., TSAI, W., LEE, K., AND CHEN, L. Integrated microblog sentiment analysis from user's social interaction patterns and textual opinions. *Applied Intelligence* 23, (2015), 1-15.

[21] LIU. B., HU, M., AND CHENG, J. Opinion observer: analyzing and comparing opinions on the Web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web* (ACM, New York, NY, USA, May 2005).

[22] LIU, Y., YU, X., AN, A., AND H, X. Riding the tide of sentiment change: sentiment analysis with evolving online reviews. *Journal World Wide Web* 16, 4 (July 2013), 477-496.

[23] LIU, B. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer-Verlag Berlin Heidelberg, 2011.

[24] MARUYAMA, O. Heterodimeric Protein Complex Identification. In *Proceedings of the 2Nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine* (Chicago, Illinois, 2011).

[25] MEDHAT, W., HASSAN, A., AND KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* 5, (2014), 1093-1113.

[26] MILLER, G., BECKWITH, R., FELLBAUM, C., GROSS, D., AND MILLER, K. Introduction to wordnet: An on-line lexical database. *International journal of lexicography* 3, 4 (1990), 235-244.

[27] MOBASHER, B., COOLEY, R., AND SRIVASTAVA, J. Automatic personalization based on web usage mining. *Communications of the ACM* 43, 8 (2000), 142-151.

[28] NIGAM, K., MCCALLUM, A., THRUN, S., AND MITCHELL, T. Text Classification from Labeled and Unlabeled Documents using EM. *Machine Learning* 39, 2 (May 2000), 103-134.

[29] PANG, B., AND LEE, L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (Barcelona, Spain, 2004).

[30] PANG, B., AND LEE, L. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval* 2, 1-2 (2008), 1-135.

[31] PATHAK, N., MANE, S., SRIVASTAVA J. AND CONTRACTOR, N.S. Knowledge Perception Analysis in Social Network. In *Proceedings of the 6th SIAM–International Conference on Data Mining* (SDM 2006).

[32] PUSTEJOVSKY, J. The generative lexicon. *Computational linguistics* 17, 4 (1991), 409-441.

[33] RASCHKA, S. Naive Bayes and Text Classification I-Introduction And Theory. *ARXIV PREPRINT ARXIV 1410, 5329* (2014).

[34] RAVINDRAN, S., AND KALPANA, M. Student's Expectation, Perception and Satisfaction towards the Management Educational Institutions. *Procedia Economics and Finance* 2, (2012), 401-410.

[35] RENNIE, J.D.M., SHIH, L., TEEVAN, J., AND KARGER, D. R. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In *International Conference on Machine Learning (*August 22, 2003).

[36] SCHNEIDER, K. A new feature selection score for multinomial naive Bayes text classification based on KL-divergence. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions* (Barcelona, Spain , July 2004).

[37] SHAN, H., AND BANERJEE, A. Mixed-membership naïve Bayes models. *Data Mining and Knowledge Discovery* 23, 1 (July 211), 1-62.

[38] SHIMODAIRA,H. Text Classification using Naïve Bayes. Retrieved August 1, 2015 from https://web.standford.edu/class/cs124/lec/naive bayes.pdf

[39] SINGH, N., JAIN, A., & RAW, R. S. Comparison Analysis Of Web Usage Mining Using Pattern Recognition Techniques. *International Journal of Data Mining & Knowledge Management Process (IJDKP)* 3, (2013).

[40] SOUZA, M., VIEIRA, R., BUSETTI, D., CHISHMAN, R., AND ALVES, I. M. Construction of a portuguese opinion lexicon from multiple resources. *STIL,* (2011).

[41] SRIVASTAVA, J., COOLEYZ, R., DESHPANDE, M., AND TAN, P.N. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter* 1, 2 (2000), 12-23.

[42] SRIVASTAVA, T., DESIKAN, P., AND KUMAR, V. Web mining–concepts, applications and research directions. *Foundations and advances in data mining,* (2005), 275-307.

[43] STARNER, T., WEAVER, J., AND PENTLAND, A. Real-time american sign language recognition using desk and wearable computer based video. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 20, 12 (1998), 1371-1375.

[44] Wang, Y., (2000), Web Mining and Knowledge Discovery of Usage Patterns. *Cs 748T Project.* (2000), 1-25.

[45] WANG, S., JIANG, L., AND LI, C. Adapting naive Bayes tree for text classification. *Knowledge and Information Systems* 44, 1 (July 2015), 77-89.

[46] WONG, T. Generalized Dirichlet priors for Naïve Bayesian classifiers with multinomial models in document classification. *Journal Data Mining and Knowledge Discovery* 28, 1 (January 2014), 123-144.